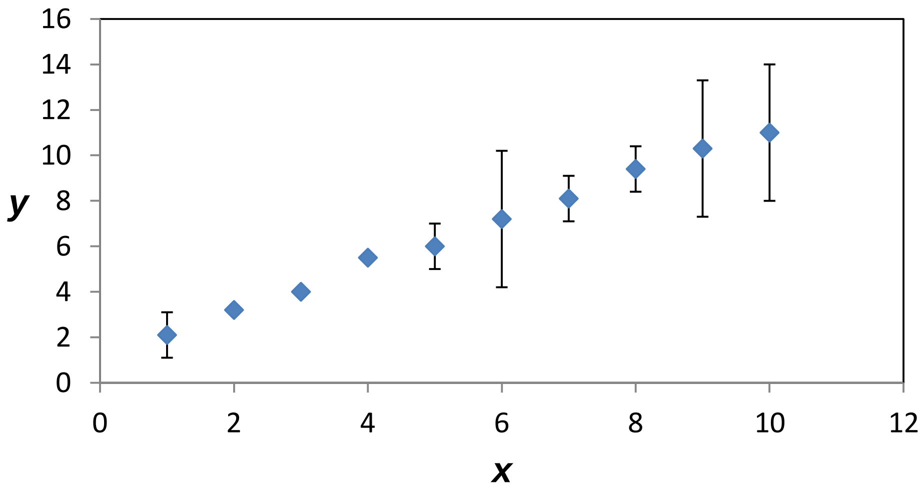
Session B: Weighted Least Squares Fitting

B.1 Introduction

In the last session we introduced the concept of "least squares" in order to obtain the best fit of a theoretical function to a set of data. In doing so, we implicitly assumed that each measurement (or data point) was equally reliable. However, in general, we do not always have data with equal measurement error. See, for example, the figure below:



It would seem sensible to take account of the variation in uncertainty in the least squares method, putting greater weight on those data with small uncertainties, less weight on those with large uncertainties. In "*weighted* least squares fitting" it is conventional to weight each measurement with the inverse of its measurement error. The weighted sum of the squares of the deviations, S, is then given by:

$$S = \sum_{i=1}^{n} \left[\frac{y_i - y_{\text{fit}}(x_i)}{\sigma_i} \right]^2$$

The y_i are measured values of y, y_{fit} are the values of y from an expression linking x and y evaluated at the measured x_i, and σ_i is the measurement error. We still seek the minimum value of S. If we consider a first order polynomial (straight line) for y_{fit} in the first instance, writing $\omega = 1/\sigma^2$, we now obtain:

$$\begin{pmatrix} [\omega x^2] & [\omega x^1] \\ [\omega x^1] & [\omega x^0] \end{pmatrix} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix}$$

or

$$A \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix} \qquad \text{where} \qquad A = \begin{pmatrix} [\omega x^2] & [\omega x^1] \\ [\omega x^1] & [\omega x^0] \end{pmatrix}$$

and

$$[\omega x^m] \quad \text{implies} \quad \sum_{i=1}^{n} \omega_i x_i^m$$

Go back and look at session A to see how we derived this last time. In brief, a first order polynomial has two (unknown) parameters, p_1 and p_0. To find the best-fit line, we differentiate S with respect to each parameter. This gives us a pair of simultaneous equations that we can solve for the two parameters. This is most conveniently expressed in matrix form.

The solution for the parameters p_1 and p_0 are obtained by inverting the matrix A:

$$\begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = A^{-1} \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix}$$

Note that the weighting implies that we shall obtain a different solution to that obtained without weighting.

The reason why it is convenient to write the simultaneous equations in matrix form is that we can exploit the symmetry of the equations to easily extend the approach to perform a least squares fit when the theoretical function is a second order (or higher) polynomial. As we saw in Session A, the symmetry of the equations in matrix form still goes through a rather straightforward progression. For a second order polynomial we obtain:

$$A \begin{pmatrix} p_2 \\ p_1 \\ p_0 \end{pmatrix} = \begin{pmatrix} [\omega x^2 y] \\ [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix} \qquad \text{where} \qquad A = \begin{pmatrix} [\omega x^4] & [\omega x^3] & [\omega x^2] \\ [\omega x^3] & [\omega x^2] & [\omega x^1] \\ [\omega x^2] & [\omega x^1] & [\omega x^0] \end{pmatrix}$$

.

The solution is then:

$$\begin{pmatrix} p_2 \\ p_1 \\ p_0 \end{pmatrix} = A^{-1} \begin{pmatrix} [\omega x^2 y] \\ [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix}$$

.

Exercise B.1

1. Take the data from the graph at the beginning of this session and do a *weighted* straight line fit using a spreadsheet.

x	y	σ_y
1	2.1	1
2	3.2	0.1
3	4	0.1
4	5.5	0.1
5	6	1
6	7.2	3
7	8.1	1
8	9.4	1
9	10.3	3
10	11	3

   Check your results against the solutions from POLYFIT (Excel's Trendline only does *unweighted* fits, so cannot be used as a comparison here).

2. Take the data from Exercise A.2. Construct standard deviations for the y values assuming that the uncertainties are given by the square root of the y value (i.e. normal counting

statistics). Do a weighted second order fit. Once again, you can check your results against the values from POLYFIT.
Note that the weighting leads to somewhat different values for the parameters compared with Exercise A.2 where we used an unweighted least squares fit.

B.2 Chi-squared

The variable, S, that we have been minimizing is usually denoted as chi-squared, χ^2. When we perform a weighted (or unweighted) least squares fit, what we are actually doing is finding the theoretical function that minimizes the value of chi-squared.

It is important to appreciate that in some circumstances we cannot find the differential of χ^2 (or S if you prefer) with respect to the parameters of the fit. Take for example a function of the form $y = p_1 \sin(p_0 x)$. When you differentiate S (i.e. χ^2) with respect to the parameters p_0 and p_1 you end up with equations that cannot be solved.

In this case the parameters cannot be found analytically. However, we can still vary p_0 and p_1 numerically, and iterate to find the values of the parameters that give the smallest value of χ^2. P. R. Bevington's book (see below) has routines for doing this. In the teaching laboratory we use this technique to solve for the cosine distribution of thermal neutrons in the graphite stack.

Exercise B.2

1. Take the data from Exercise B.1 and select values of p_1 and p_0 that are near but not equal to the best-fit values you found. Calculate the corresponding χ^2.
2. Now adjust one of the parameters, either p_1 or p_0 (it does not matter which you vary first) and recalculate. Continue the process until the value of χ^2 starts to rise again. Now, using the value of the selected parameter, which has given the minimum χ^2, perform the same process on the other parameter.
3. Iterate until the lowest possible value of χ^2 has been obtained.
4. Compare this with the results from Exercise B.1

In fitting problems involving functions other than polynomials, where the solution is *not* analytic, an iterative procedure must be used. The simplest iterative method is one that varies each parameter in turn in order to locate a *local* minimum in χ^2. This will not be the lowest possible value of χ^2, unless the other parameter(s) are also at their optimal values. Several iterations will be required to locate the true global minimum and hence find the best-fit parameters in this case. This kind of procedure is best done by a computer program, rather than using a spreadsheet.

Further Reading

P. R. Bevington "Data Reduction and Error Analysis for the Physical Sciences", McGraw Hill (1969)